vEGU21 ITS4.4/AS4.1

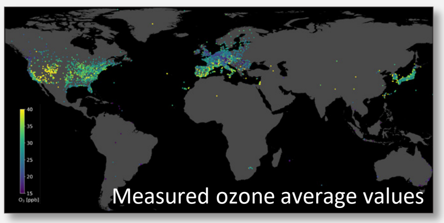
Global fine resolution mapping of ozone metrics through explainable machine learning

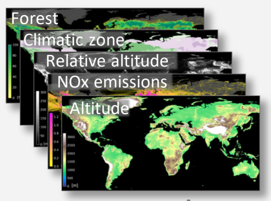
Clara Betancourt (c.betancourt@fz-juelich.de), Scarlet Stadtler, Timo Stomberg, Ann-Kathrin Edrich, Ankit Patnala, Ribana Roscher, Julia Kowalski, and Martin G. Schultz

View abstract

Data description

A-priori uncertainty estimation





\+ ...

We combine global air quality measurements with high-resolution geospatial data from various sources.

Algorithms, configuration

Scientific consistency

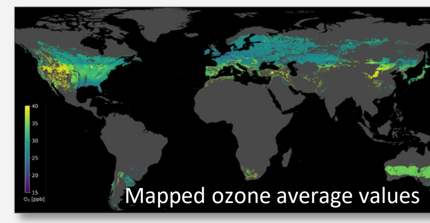
Error modeling

Model robustness

Explainable machine learning is applied to obtain estimates of the ozone concentration based on the geospatial data.

Area of applicability

Map interpretation



The learned model is used to obtain a global high-resolution air quality map, including locations without measurements.

Summary + outlook

This is a nonlinear presentation. Please use the buttons to click through the content. Thanks!

Back to title slide

c.betancourt@fz-juelich.de

Abstract

Back to title slide

Global fine resolution mapping of ozone metrics through explainable machine learning

Through the availability of multi-year ground based ozone observations on a global scale, substantial geospatial meta data, and high performance computing capacities, it is now possible to use machine learning for a **global data-driven ozone assessment**. In this presentation, we will show a novel, completely data-driven approach to map tropospheric ozone globally.

Our goal is to interpolate ozone metrics and aggregated statistics from the database of the Tropospheric Ozone Assessment Report (TOAR) onto a global 0.1° x 0.1° resolution grid. It is challenging to interpolate ozone, a toxic greenhouse gas because its formation depends on **many interconnected environmental factors** on small scales. We conduct the interpolation with various machine learning methods trained on aggregated hourly ozone data from five years at more than 5500 locations worldwide. We use several geospatial datasets as training inputs to provide proxy input for environmental factors controlling ozone formation, such as precursor emissions and climate. The resulting maps contain different ozone metrics, i.e. statistical aggregations which are widely used to assess air pollution impacts on health, vegetation, and climate.

The key aspects of this contribution are twofold: First, we apply **explainable machine learning** methods to the data-driven ozone assessment. Second, we discuss dominant **uncertainties** relevant to the ozone mapping and quantify their impact whenever possible. Our methods include a thorough a-priori uncertainty estimation of the various data and methods, assessment of scientific consistency, finding critical model parameters, using ensemble methods, and performing error modeling.

Our work aims to increase the reliability and integrity of the derived ozone maps through the provision of scientific robustness to a data-centric machine learning task. This study hence represents a blueprint for how to formulate an environmental machine learning task scientifically, gather the necessary data, and develop a data-driven workflow that focuses on optimizing transparency and applicability of its product to **maximize its scientific knowledge return**.

Doi: https://doi.org/10.5194/egusphere-egu21-7596

c.betancourt@fz-juelich.de

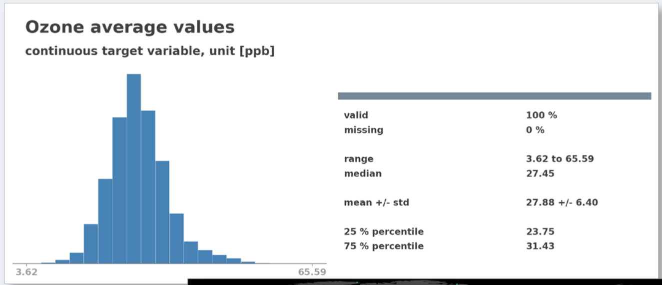
Back to title slide

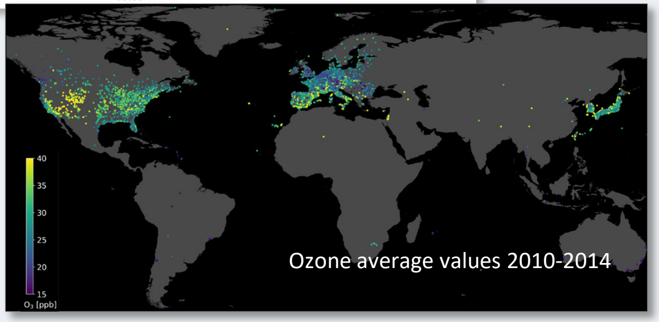
Data description

For our mapping approach, we combine multi-year ground based ozone observations on a global scale with several high-resolution geospatial datasets.

Ozone data

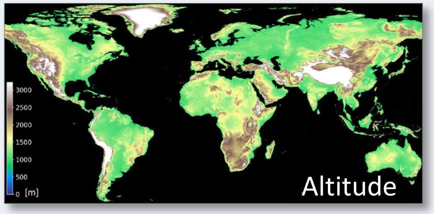
- We use the AQ-Bench dataset as training data
- It contains aggregated ozone metrics from the years 2010-2014 at 5577 air quality observation stations of the TOAR database
- Ozone metrics include: average ozone, percentiles, health/ vegetation related metrics. *(This display only contains maps of average ozone)*
- Most stations are located in the US, Europe or East Asia (≈ 98 %)

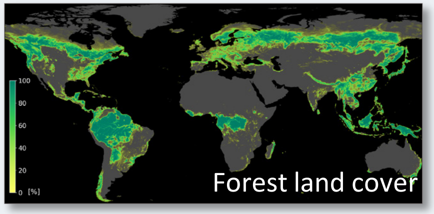


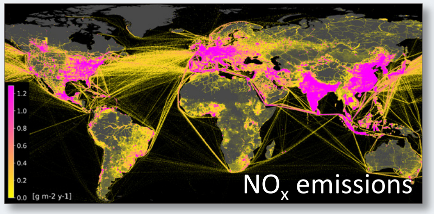


Geospatial data

- The chosen geospatial data are related to ozone processes, as described in Betancourt et al. 2020
- The data are gathered from multiple sources and harmonized to a 0.1° × 0.1° resolution
- All data fields: Climatic zone, altitude/relative altitude, several types of land cover, wheat/rice production, NO_x emissions, NO_2 full column, population density, stable nightlights, latitude







+ ...

c.betancourt@fz-juelich.de

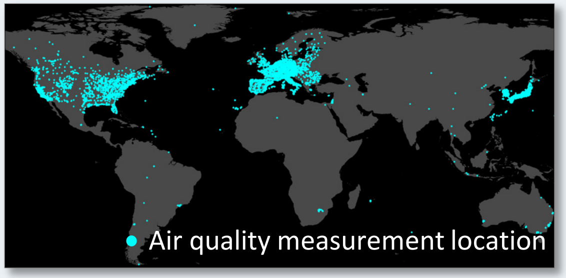
A-priori uncertainty estimation

Before applying machine learning, we evaluate the data and planned mapping method. We also formulate the expected accuracy of our mapping method (see grey box).

Back to title slide

Ozone data uncertainty

- Ozone measurement error: 5 ppb is a conservative estimate (TOAR-db)
- Robustness of ozone metrics is ensured by data capture criteria
- Ozone trends are usually < 1ppb/y
- Measurements spread irregularly over the globe. Most measurements in USA, Europe, Japan (see image)



Geospatial data uncertainty

- The geospatial data are collected from multiple sources, such as satellite images and digital elevation models
- All sources have their own quality control
- Through global change, also the geospatial data can show trends

Modeling approach issues

- Training data coverage is low in some world regions, e.g. Africa
- We use a snapshot approach where temporal variations are filtered out
- The model robustness might be an issue in some regions since we train on a small dataset of 5577 samples
- Point measurements might not be representative for 0.1° × 0.1° grid boxes

We expect the model to reliably capture the global variability of ozone in all regions with sufficient training data coverage.

c.betancourt@fz-juelich.de

Machine learning algorithms and configuration

Back to title slide

Our dataset is relatively small, so we use standard machine learning methods (shallow Neural Network and Random Forest), and no deep neural networks. These methods are quick to train and explainable machine learning methods can be applied relatively easily.

Algorithms and training

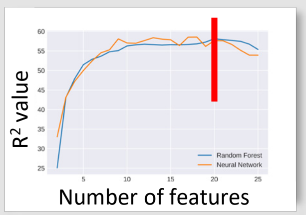
- 60/20/20% independent data split
  (stations are independent if > 50 km apart)
- Neural Network with two hidden layers (20/5 neurons)
  - Learning rate = 0.0001
  - λ_2 (regularization parameter) = 0.05
  - ReLu activation function
  - 15000 training epochs with an Adam optimizer
- Random Forest with 100 trees

Feature Engineering

- We performed basic feature engineering to improve the interpretability of the model
- Different types of savanna, shrublands, and forests are given individually in the original land cover dataset. We merged them into 'savanna', 'forest' and 'shrubland'
- Instead of the latitude, we took the absolute latitude, since radiation and temperature decrease when moving away from the equator regardless in which direction

Pruning

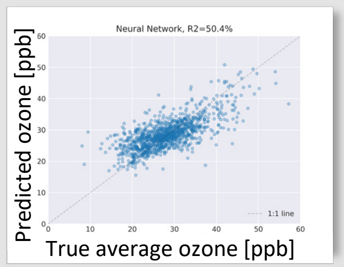
- Unnecessary predictors can favor overfitting
- Five predictors decreased the test R^2 value of both algorithms, so we removed them

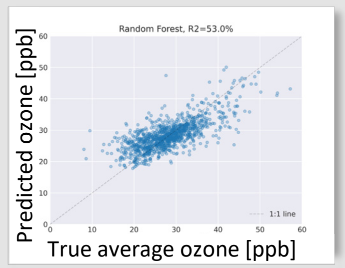


Final predictor list

Evaluation metric

- R^2 > 50% for both algorithms on independent test set





Our model performs well on the independent test set.

c.betancourt@fz-juelich.de

Final predictor list

Algorithms, configuration

Climatic zone
Absolute latitude
Altitude
Relative altitude
Water in 25 km area
Forest in 25 km area
Shrublands in 25 km area
Savannas in 25 km area
Grasslands in 25 km area
Permanent wetlands in 25 km area
Croplands in 25 km area
Rice production
NO_x emissions
NO_2 column
Population density
Maximum population density in 5 km area
Maximum population density 25 km area
Nightlight 1 km
Nightlight in 5 km area
Maximum nightlight in 25 km area

Urban And Built-Up in 25 km area
Cropland / Natural vegetation mosaic in 25 km area
Snow and ice in 25 km area
Barren or sparsely vegetated in 25 km area
Wheat production

pruned

c.betancourt@fz-juelich.de

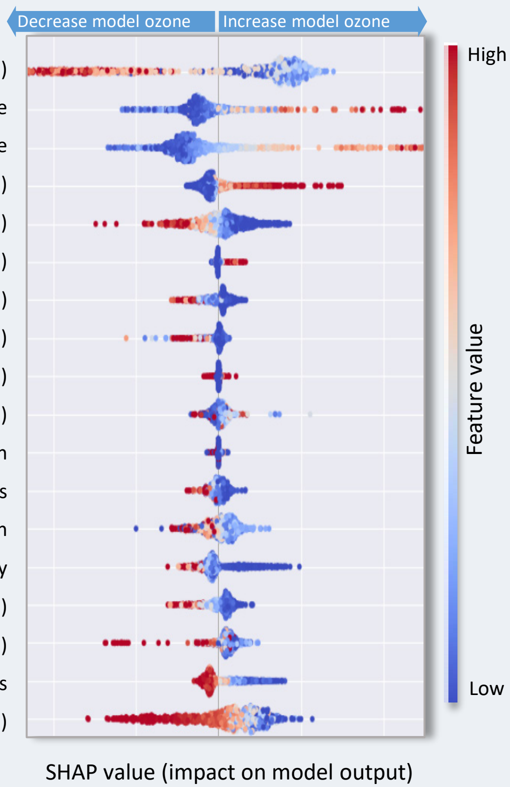
Scientific consistency

Back to title slide

We use SHAP values (short for SHapley Additive exPlanations) to evaluate the scientific consistency of our model.
A SHAP value indicates the influence a feature has on the model result and how large this influence is.

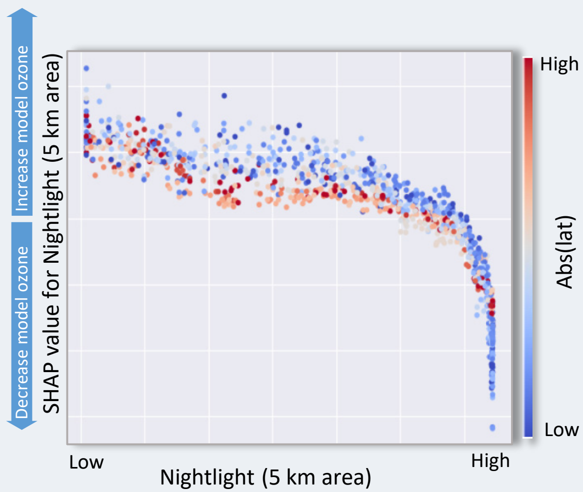
Feature importance

The graph below shows some SHAP values of our random forest model on the test set. One example is the latitude, a proxy for warmth and radiation, which favors ozone formation. The SHAP value increases when moving closer to the equator, which is **consistent** with ozone research.



Feature interaction

SHAP values also show the interaction of different features in the model. The plot below shows that the influence of Nightlight (a proxy for industry and traffic exhaust) on the model results is generally nonlinear and also depends on warmth and radiation (proxied by latitude).



The model captures nonlinear relations which are consistent with our understanding of ozone chemistry

IntelliAQ KI:STE RWTH AACHEN UNIVERSITY UNIVERSITÄT BONN JÜLICH Forschungszentrum

c.betancourt@fz-juelich.de

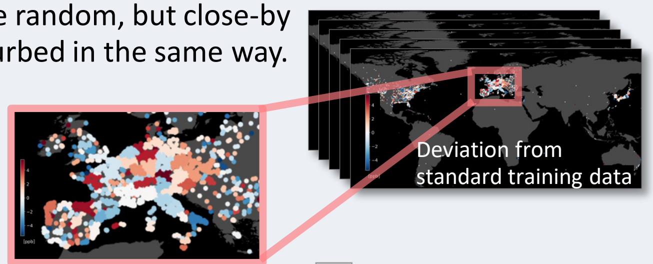
Error modeling

Our model is expected to have errors and uncertainties. Here we assess the error introduced into our model by spatial fluctuations in ozone values. A typical spatial ozone fluctuation is 5 parts per billion (TOAR-db). But just perturbing the training data with random noise of that amplitude would ignore the correlation of the spatial fluctuations, and may lead to an underestimation of the introduced error. Instead, stations which are close to each other (< 50 km apart) are perturbed similarly.

Back to title slide

Spatial perturbations

1) Error model for ozone measurements: **perturbations** are random, but close-by stations are perturbed in the same way.

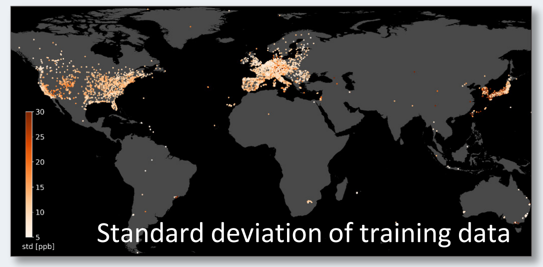


2) The model is **retrained** on the perturbed inputs

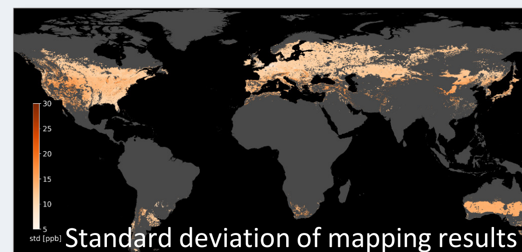
3) A map is generated based on the perturbed model. It shows that the **deviation** of our standard map never exceeds the initial perturbation of the training data.

Influence on model results

4) **Aggregate** results: Multiple realizations of spatial perturbations provide an ensemble of models. Looking at the standard deviations, we see that ozone fluctuations introduce a globally relatively uniform error which is in the range of the fluctuations.



The model is robust against typical ozone fluctuations.



5) Another error model was used to investigate **temporal fluctuations** and trends in the ozone. It showed similar results.

IntelliAQ KI:STE RWTH AACHEN UNIVERSITY UNIVERSITÄT BONN JÜLICH Forschungszentrum
c.betancourt@fz-juelich.de

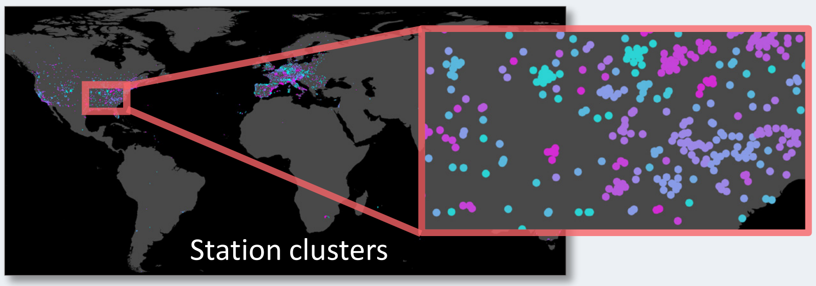
Model robustness

Regarding the robustness of the model, two things need to be investigated: 1) Is the model itself robust? 2) Can the model be applied in all world regions with similar features to the training dataset, or is it spatially overfitted?

Back to title slide

Robustness analysis by cross validation

1) To evaluate the robustness of the model, we group the stations into interdependent clusters, and then assign those clusters to training/evaluation sets randomly (the independent test set is left out completely).

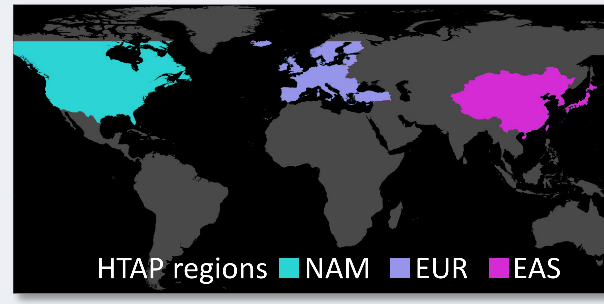


2) From the groups of independent stations, 4 independent datasets are formed, which are then used for cross validation.

3) Although the test R^2 value was stable, the training and validation R^2 values vary by 7-8% for both Random Forest and Neural Network. This was expected, because with 5577 samples, the dataset is relatively small.

Spatial overfitting

1) To test spatial overfitting, we divide the world into regions. >98% of our training data are in either in NAM, EUR or EAS.



2) Then we train our model only on two of the three world regions (e.g. EUR and EAS).

3) Testing on the remaining region (NAM), the R^2 value of the random forest model is diminished by 10%, and that of the neural network drops by 40%. So the random forest seems to be less prone to spatial overfitting. More investigation is needed on this issue.

The dataset is slightly noisy. Furthermore, the issue of spatial overfitting has to be investigated more.

IntelliAQ
KI:STE
RWTH AACHEN UNIVERSITY
UNIVERSITÄT BONN
JÜLICH Forschungszentrum
c.betancourt@fz-juelich.de

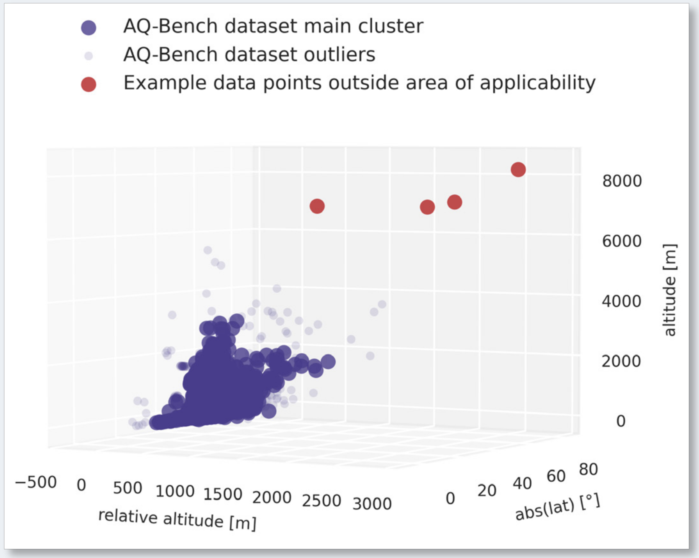
Area of applicability

Back to title slide

Our model is technically applicable on the whole world. But there are regions which have properties very different from our training data. It is important to sort them out as the predictions are not valid there. Our 'Area of Applicability' is based on a preprint by Meyer & Pebezma, ArXiv, 2020.

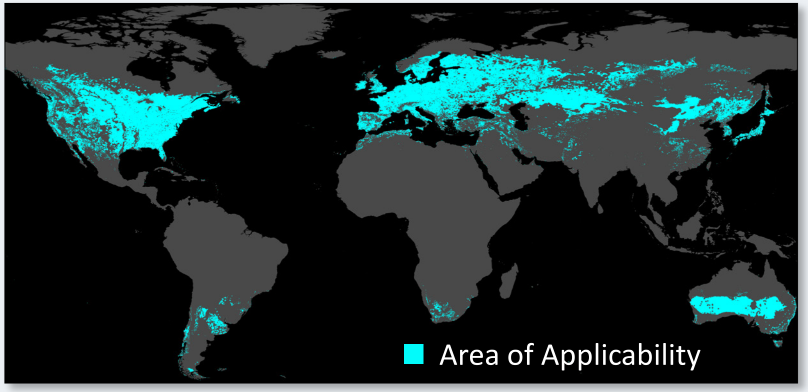
Training data form cluster in feature space

1) Our training data set AQ-Bench forms a cluster in the feature space (cluster found with dbscan). Before the ozone at a specific location is predicted, we check if its features are part on this cluster. The plot below shows a 3D projection of the feature space. Example data points we excluded from mapping are shown in red.



We map only the points inside that cluster

2) The Area of Applicability covers most parts of USA, Europe and Asia. Environmental conditions often differ from the training data in the rest of the world. High mountains are not in the Area of Applicability, because there are hardly any ozone measurement stations built at higher altitudes.



We only apply the model in regions with features values similar to those in the training dataset.

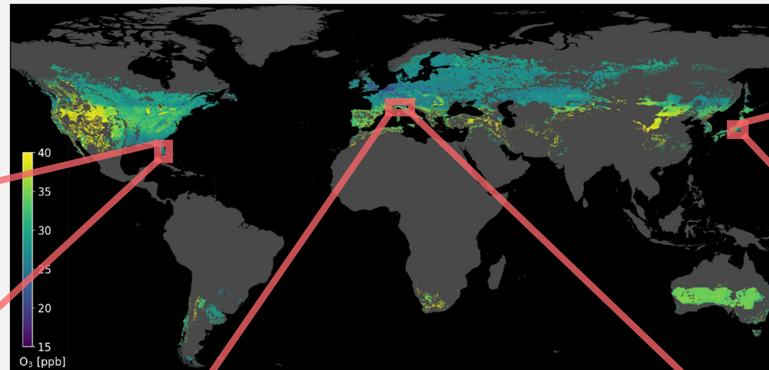
IntelliAQ KI:STE RWTH AACHEN UNIVERSITY UNIVERSITÄT BONN JÜLICH Forschungszentrum

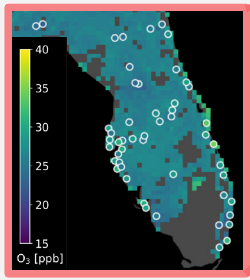
c.betancourt@fz-juelich.de

Map interpretation

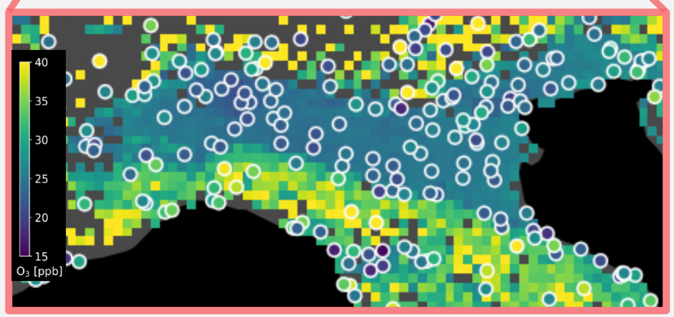
Back to title slide

Below you can see our map generated with the Random Forest model. We have added a detailed view of some exemplary regions with interesting ozone patterns.

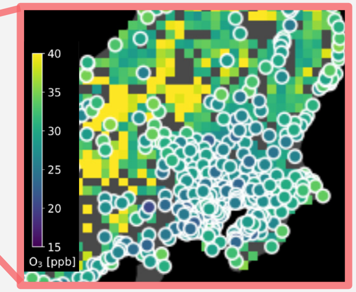




Swamp/coast (Florida):
We see a relatively uniform distribution, with slightly elevated ozone values at the eastern coast.



Po valley and Alps (Italy): The Po valley is located close to the alps in northern Italy. We can see clearly that low ozone is captured in the valley and rises at mountains, as we expected from previous studies.



Tokyo and surroundings (Japan):
The metropolitan region around Tokyo is well covered with measurement stations.
Predictions become more important outside the area.

True values

Our model

Spatial patterns
(e.g. coast, mountain, city)
are captured well

IntelliAQ KI:STE RWTH AACHEN UNIVERSITY UNIVERSITÄT BONN JÜLICH Forschungszentrum

c.betancourt@fz-juelich.de

Summary and outlook

Back to title slide

- ✓ We produced the first data driven map of global tropospheric ozone, $R^2 > 50\%$
- ✓ Mapping allows ozone assessment in regions with no air quality stations, but with similar environmental conditions as in the training data
- ✓ Explainable machine learning increases trust in the maps

Future research…

- ➢ More air quality observations are needed to extend the area of applicability, e.g. in Africa or South America
- ➢ The maps could be compared with ozone model data
- ➢ Time resolved mapping can be conducted, instead of our ‘snapshot’ approach

References

Authors of this display:

Clara Betancourt (c.betancourt@fz-juelich.de, info), Scarlet Stadtler (info), Timo Stomberg (info), Ann-Kathrin Edrich (info), Ankit Patnala (info) Ribana Roscher (info), Julia Kowalski (info), and Martin G. Schultz (info)

Projects:

IntelliAQ (info), KISTE (info), TOAR (link)

References:

Betancourt et al. “AQ-Bench: A Benchmark Dataset for Machine Learning on Global Air Quality Metrics”, ESSD discussions 2021 (link)
Lundberg and Lee “A Unified Approach to Interpreting Model Predictions”, arXiv 2017 (link)
Meyer and Pebezma “Predicting into unknown space? Estimating the area of applicability of spatial prediction models” , arXiv 2020 (link)
Schultz et al. “Tropospheric Ozone Assessment Report: Database and Metrics Data of Global Surface Ozone Observations”, Elementa 2017 (link)

c.betancourt@fz-juelich.de